ARE YOU ABLE TO TELL STORIES WITH DATA?

F. Ramos[1], S. Trilles[1,2], C. Granell[1]

[1]Universitat Jaume I (SPAIN)
[2]Universidad Isabel I (SPAIN)

Abstract

During our first education years, at the school, even at the high school, we learn very much about many different subjects. However, in most cases, an important ability for our professional future is missed: how to create and tell stories with data and numbers. This lack of information has produced a big problem: users have access to a large amount of information, due to the current technological advances, but they are unable to efficiently use data to tell stories, which is key to convert them into relevant information.

In this paper, we present a study we performed with master students aimed at improving their data visualization skills. We divided this study into several stages. Firstly, the students learnt by accomplishing different tasks and following a clear path divided into these sections: analysis of the audience, selecting the appropriate visuals, simplification, focus, communication with data, storytelling, and final visuals.

Every task was presented in a visual and natural way, with different options, where the students should choose the correct answers, from a clear and objective point of view. All the tasks are part of a global questionnaire, which was carried out by the students so that they prove their competence in telling effective stories with numbers and data.

The results shown clear lacks in different tasks such a simplification and communication with data, where most of students didn't correctly answer the questions.

As a conclusion, we can claim that more effort must be done in setting a clear objective in the visual communication field. In particular, simplify data, focus on the important part of the message to the audience, and also in the way and shape that data are presented.

Keyword: Data, visualisation, storytelling, storyboard.

1 INTRODUCTION

In general, the human cognitive system is bound to our ability to see and understand the data. It is due to the evolution that we have a special organ (our brain), connected to our eyes, which enable us to see and understand our environment. However, sometimes the system can fail. In Figure 1, we can observe the "impossible staircase", where you can have the illusion of ascending or descending yet form a continuous loop, so that a person could climb them forever and never get any higher.

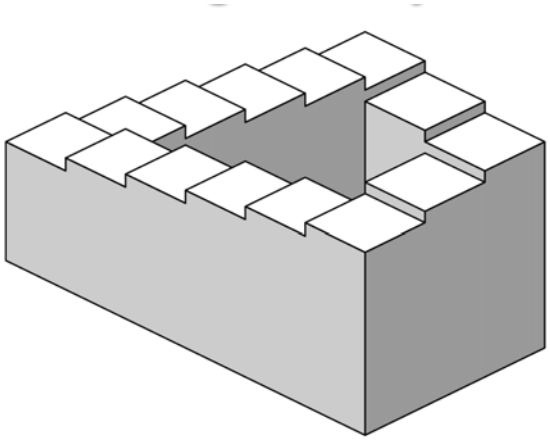

Figure 1. Impossible Staircase.

In [1], Francis Anscombe claimed that It's essential to create data visualisations before the data analysis, making both calculations and graphs. Both sorts of output should be studied and each will contribute to understanding. Thus, in the Datasaurus Dozen Project [2], it is studied how summary statistics alone couldn't not show the clear differences among some existing datasets [3]. However, the visualization of them allows us to perceive their different nature[4]. In Figure 2, it is possible to observe nice datasets that while different in appearance, each dataset has the same summary statistics. However, in this example it is normally easy for the humans to detect the difference adding the visual component.

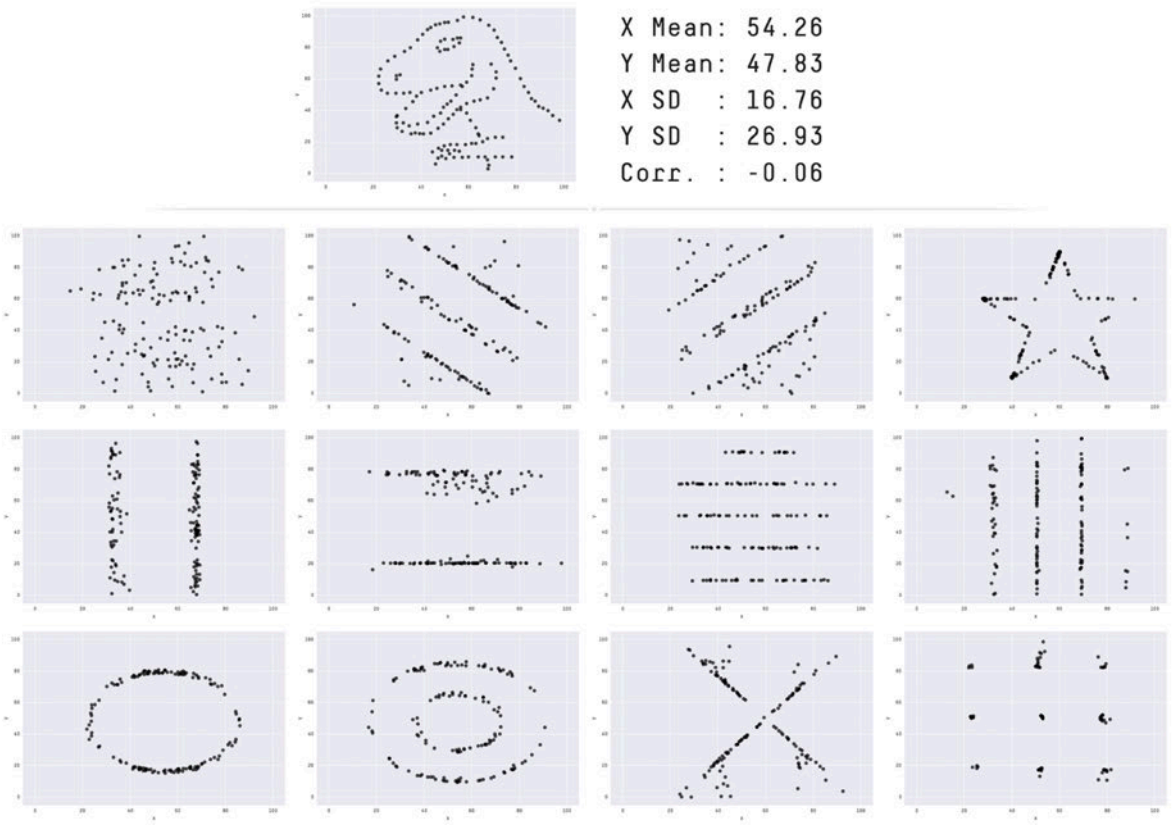

Figure 2. Nice datasets with the same statistical features but different visualizations [2].

In this context, it is also important to notice that sometimes the visualization does not help us to understand the difference among datasets. In Figure 3, some datasets, with the same summary statistics and not clearly visually distinct, are shown.

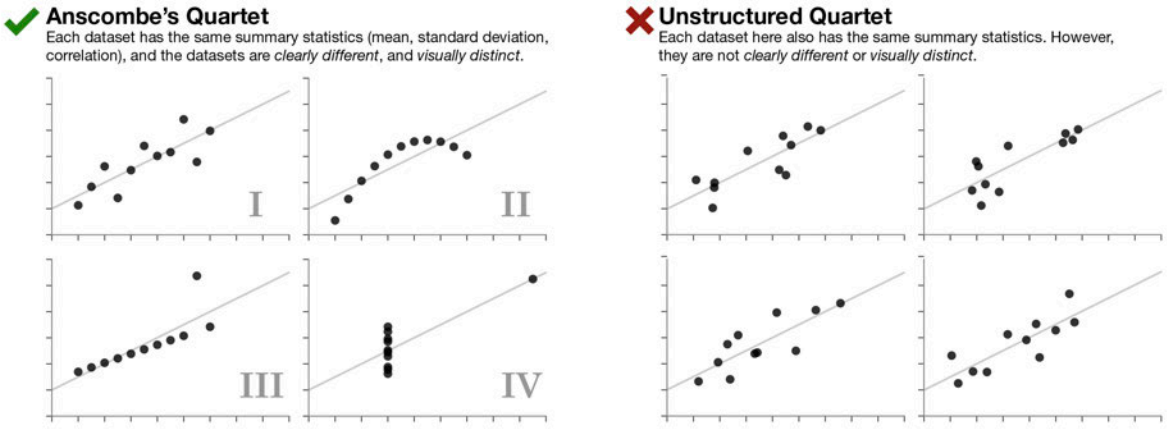

Figure 3. Datasets with the same summary statistics [2].

In general, it seems clear that human intuition matters, and humans can perceive patterns that are very difficult to be detected by computers. In Figure 4, it takes seconds to perceive different patterns and trends in the datasets with simple visualizations, which are much more difficult to detect by analysing just the numbers.

Category	2013 Q1	2013 Q2	2013 Q3	2013 Q4	2014 Q1	2014 Q2	2014 Q3	2014 Q4
Furniture	$463,988	$352,779	$338,169	$317,735	$320,875	$287,934	$319,537	$324,319
Office Supplies	$232,558	$290,055	$265,083	$246,946	$219,514	$202,412	$198,268	$279,679
Technology	$563,866	$244,045	$432,299	$461,616	$285,527	$353,237	$338,360	$420,018
Category	**2015 Q1**	**2015 Q2**	**2015 Q3**	**2015 Q4**	**2016 Q1**	**2016 Q2**	**2016 Q3**	**2016 Q4**
Furniture	$307,028	$273,836	$290,886	$397,912	$337,299	$245,445	$286,972	$313,878
Office Supplies	$207,363	$183,631	$191,405	$217,950	$241,281	$286,548	$217,198	$272,870
Technology	$333,002	$291,116	$356,243	$386,445	$386,387	$397,201	$359,656	$375,229

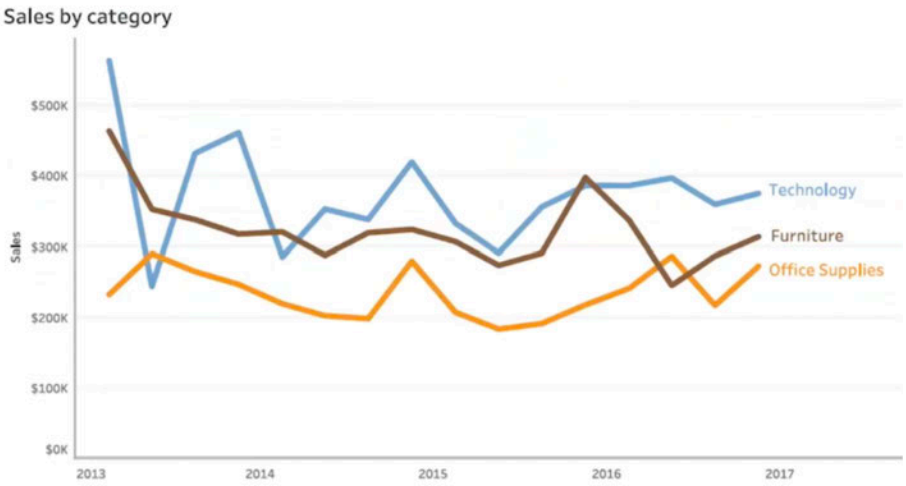

Figure 4. Datasets, with simple data visualizations (line charts), show clear patterns [5].

2 METHODOLOGY AND RESULTS

In our work, we passed some tests to our students aimed at improving their data visualization skills and at showing them how visualization can help them to better understand the data and convert them into information. Before the test, they were informed about the different ways to visualize data and how to represent them.

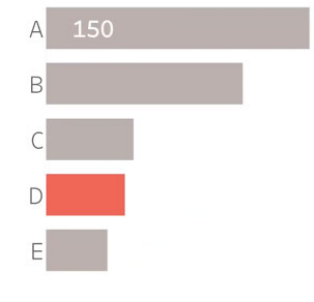

Figure 5. Test 1. What is the value of D?

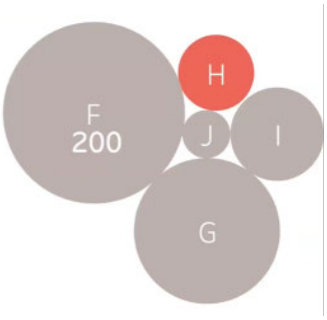

Figure 6. Test 2. What is the value of H?

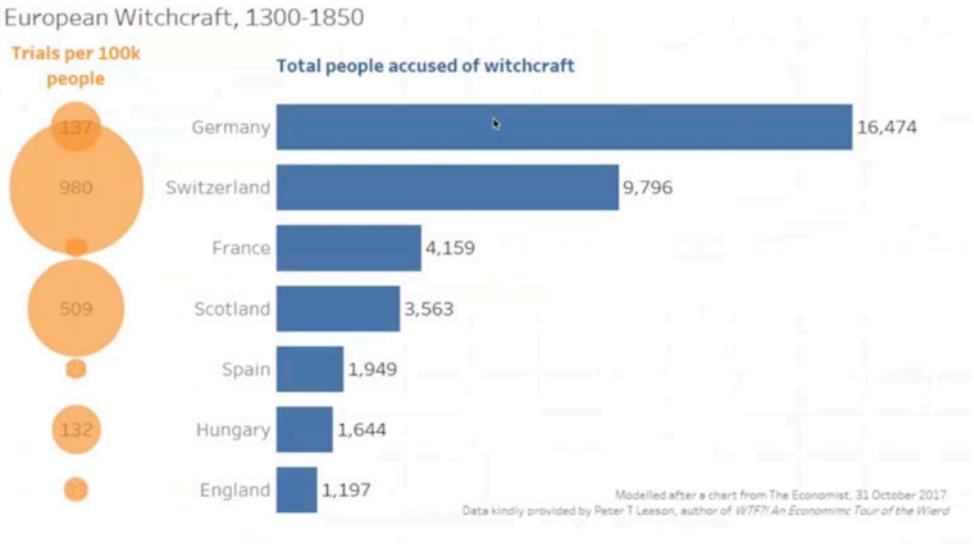

Figure 7. Test 3. What is the information extracted watching this chart?

In Figure 5, 6 and 7, we show the questions answered by the students in our study, which were presented to them during a short period of time. In this case, students where much closer to figure out the real value of D (45) than to H (35), proving that, in this example, a visualization using “lengths” is much better than “sizes”. In Figure 8, it is possible to observe the results from our test.

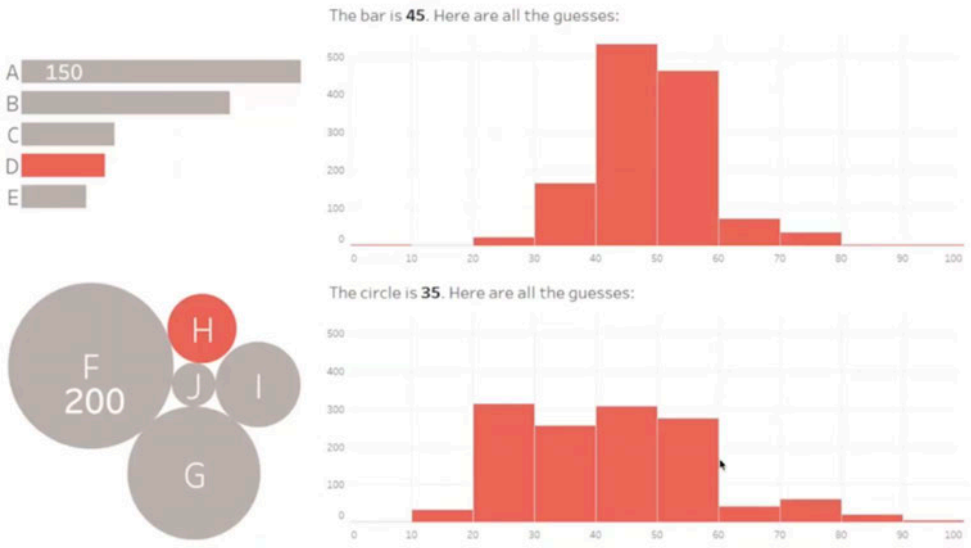

Figure 8. Results from Test1 and Test2.

Therefore, we followed a methodology based on visual questionnaires by using an online voting system that write all the information into a database. After that, we analysed the data and computed a set of statistics that support our conclusions.

3 CONCLUSIONS

In this work, we presented a preliminary study, which it is still in progress, aimed at underlining the importance of the visual component in the representation and understanding of different datasets. We discussed the different and more convenient ways to visually represent information and why the nature of the data and the message to transmit are strongly bounded to the type of visualization to be used.

Indeed, as previously commented, we created an online voting system to show our test to a high number of students. Results show that humans are better in detecting changes in length than in size, from the unconscious attributes (Figure 9), which are inherent to the human conduction. However, the combination of both can also give important insights used appropriately.

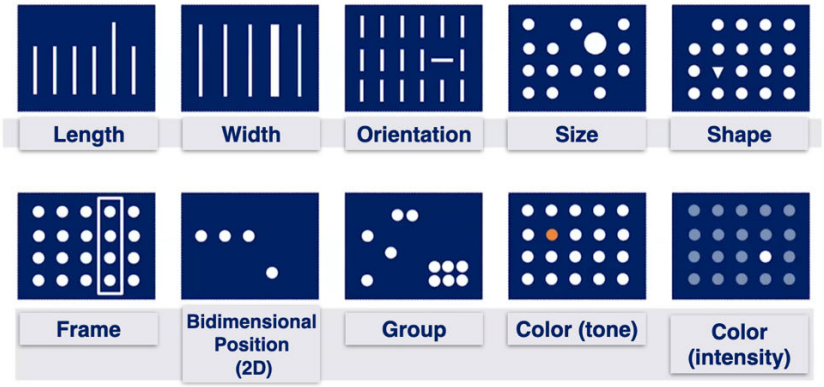

Figure 9. Unconscious attributes detected effortless by humans..

ACKNOWLEDGEMENTS

This research is partially funded by the AICO 2023 project (grant number CIAICO/2022/111) of the Department of Innovation, Universities, Science and Digital Society of the Valencian Government, Spain. and by the Research Promotion Plan of the Universitat Jaume I of Castelló, Spain (UJI-B2022-64).

REFERENCES

[1] F. J. Anscombe . Graphs in Statistical Analysis. The American Statistician, Vol. 27, No. 1. (Feb., 1973), pp. 17-21.

[2] Justin Matejka, George Fitzmaurice. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. (May 2017). pp. 1290–1294.

[3] Bach, B., Spritzer, A., Lutton, E., and Fekete, J.-D. (2012). Interactive Random Graph Generation with Evolutionary Algorithms. *SpringerLink*, 541–552.

[4] Chatterjee, S. and Firat, A. (2007). Generating Data with Identical Statistics but Dissimilar Graphics. *The American Statistician 61*, 3, 248–254.

[5] Tableau Software. Url: https://www.tableau.com